Review of "Artificial Intelligence: A General Survey"

John McCarthy
Computer Science Department
Stanford University
Stanford, CA 94305
Internet: jmc@cs.stanford.edu

June 13, 2000

Review of "Artificial Intelligence: A General Survey" by Professor Sir James Lighthill, FRS, in *Artificial Intelligence: a paper symposium*, Science Research Council 1973.

Professor Lighthill of Cambridge University is a famous hydrodynamicist with a recent interest in applications to biology. His review of artificial intelligence was at the request of Brian Flowers, then head of the Science Research Council of Great Britain, the main funding body for British university scientific research. Its purpose was to help the Science Research Council decide requests for support of work in AI. Lighthill claims no previous acquaintance with the field, but refers to a large number of authors whose works he consulted, though not to any specific papers.

The *Lighthill Report* is organized around a classification of AI research into three categories:

Category A is *advanced automation* or *applications*, and he approves of it in principle. Included in A are some activities that are obviously applied but also activities like computer chess playing that are often done not for themselves but in order to study the structure of intelligent behavior.

Category C comprises studies of the *central nervous system* including computer modeling in support of both neurophysiology and psychology.

Category B is defined as “building robots” and “bridge” between the other two categories. Lighthill defines a robot as a program or device built neither to serve a useful purpose nor to study the central nervous system, which obviously would exclude Unimates, etc. which are generally referred to as industrial robots. Emphasizing the bridge aspect of the definition, Lighthill states as obvious that work in category B is worthwhile only in so far as it contributes to the other categories.

If we take this categorization seriously, then most AI researchers lose intellectual contact with Lighthill immediately, because his three categories have no place for what is or should be our main scientific activity - *studying the structure of information and the structure of problem solving processes independently of applications and independently of its realization in animals or humans.* This study is based on the following ideas:

1. Intellectual activity takes place in a world that has a certain physical and intellectual structure: Physical objects exist, move about, are created and destroyed. Actions that may be performed have effects that are partially known. Entities with goals have available to them certain information about this world. Some of this information may be built in, and some arises from observation, from

communication, from reasoning, and by more or less complex processes of retrieval from information bases. Much of this structure is common to the intellectual position of animals, people, and machines which we may design, e.g. the effects of physical actions on material objects and also the information that may be obtained about these objects by vision. The general structure of the intellectual world is far from understood, and it is often quite difficult to decide how to represent effectively the information available about a quite limited domain of action even when we are quite willing to treat a particular problem in an *ad hoc* way.

2. The processes of problem solving depend on the class of problems being solved more than on the solver. Thus playing chess seems to require look-ahead whether the apparatus is made of neurons or transistors. Isolation of the information relevant to a problem from the totality of previous experience is required whether the solver is man or machine, and so is the ability to divide a problem into weakly connected subproblems that can be thought about separately before the results are combined.

3. Experiment is useful in determining what representations of information and what problem solving processes are needed to solve a given class of problems. We can illustrate this point by an example from the *Lighthill Re-*

port which asserts (p. 15) that the heuristics of a chess program are embodied in the evaluation function. This is plausible and was assumed by the first writers of chess programs. Experiment showed, however, that the procedures that select what part of the move tree is examined are even more important, i.e. when the program errs it is usually because it didn't examine a line of play rather than because it mis-evaluated a final position. Modern chess programs concentrate on this and often have simpler evaluators than the earlier programs.

4. The experimental domain should be chosen to test the adequacy of representations of information and of problem solving mechanisms. Thus chess has contributed much to the study of tree search; one Soviet computer scientist refers to chess as the *Drosophila* of artificial intelligence. I think there is much more to be learned from chess, because master level play will require more than just improving the present methods of searching trees. Namely, it will require the ability to identify, represent, and recognize the patterns of position and play that correspond to "chess ideas", the ability to solve some abstractions of positions (e.g. how to make use of a passed pawn and a seventh rank rook jointly) and to apply the result to actual positions. It will probably also require the ability to analyze a problem into subproblems and combine the separate results. (This

ability is certainly required for a successful *Go* program).

Having ignored the possibility that AI has goals of its own, Lighthill goes on to document his claim that it has not contributed to applications or to psychology and physiology. He exaggerates a bit here, it seems worthwhile to spend some effort disputing his claims that AI has not contributed to these other subjects.

In my opinion, AI's contribution to practical applications has been significant but so far mostly peripheral to the central ideas and problems of AI. Thus the LISP language for symbolic computing was developed for AI use, but has had applications to symbolic computations in other areas, e.g. physics. Moreover, some ideas from LISP such as conditional expressions and recursive function definitions have been used in other programming languages. However, the ideas that have been applied elsewhere don't have a specifically AI character and might have been but weren't developed without AI in mind. Other examples include time-sharing, the first proposals for which had AI motivations and some techniques of picture processing that were first developed in AI laboratories and have been used elsewhere. Even the current work in automatic assembly using vision might have been developed without AI in mind. However, the Dendral work has always had a specifically AI character, and many of the recent developments in programming such as PLANNER and CONNIVER have an AI motivation.

AI's contributions to neurophysiology have been small and mostly of a negative character, i.e. showing that certain mechanisms that neurophysiologists propose are not well defined or inadequate to carry out the behavior they are supposed to account for. I have in mind Hebb's proposals in his book *The Organization of Behavior*. No-one today would believe that the gaps in those ideas could be filled without adding something much larger than the original work. Moreover, the last 20 years experience in programming machines to learn and solve problems makes it implausible that cell assemblies *per se* would learn much without putting in some additional organization, and physiologists today would be unlikely to propose such a theory. However, merely showing that some things are unlikely to work is not a *positive* contribution. I think there will be more interaction between AI and neurophysiology as soon as the neurophysiologists are in a position to compare information processing models of higher level functions with physiological data. There is little contact at the nerve cell level, because, as Minsky showed in his PhD dissertation in 1954, almost any of the proposed models of the neuron is a universal computing element, so that there is no connection between the structure of the neuron and what higher level processes are possible.

On the other hand, the effects of artificial intelligence research on psychology have been larger as attested by various psychologists. First of all, psychologists have begun to use models in which complex internal data structures that cannot be observed directly are attributed to animals and people. Psychologists have come to use these models, because they exhibit behavior that cannot be exhibited by models conforming to the tenets of behaviorism which essentially allows only connections between externally observable variables. Information processing models in psychology have also induced dissatisfaction with psychoanalytic and related theories of emotional behavior. Namely, these information processing models of emotional states can yield predictions that can be compared with experiment or experience in a more definite way than can the vague models of psychoanalysis and its offspring.

Contributions of AI to psychology are further discussed in the paper *Some Comments on the Lighthill Report* by N. S. Sutherland which was included in the same book with the Lighthill report itself.

Systematic comment on the main section, entitled *Past Disappointments* is difficult because of the strange way the subject is divided up but here are some remarks:

1. Automatic landing systems for airplanes are offered

as a field in which conventional engineering techniques have been more successful than AI methods. Indeed, no-one would advocate applying the scene analysis or tree search techniques developed in AI research to automatic landing in the context in which automatic landing has been developed. Namely, radio signals are available to determine the precise position of the airplane in relation to a straight runway which is guaranteed clear of interfering objects. AI techniques would be necessary to make a system capable of landing on an unprepared dirt strip with no radio aids which had to be located and distinguished from roads visually and which might have cows or potholes or muddy places on it. The problem of automatically driving an automobile in an uncontrolled environment is even more difficult and will definitely require AI techniques, which, however, are not nearly ready for a full solution of such a difficult problem.

2. Lighthill is disappointed that detailed knowledge of subject matter has to be put in if programs are to be successful in theorem proving, interpreting mass spectra, and game playing. He uses the word *heuristics* in a non-standard way for this. He misses the fact that there are great difficulties in finding ways of representing knowledge of the world in computer programs and much AI research and internal controversy are directed to this

problem. Moreover, most AI researchers feel that more progress on this *representation problem* is essential before substantial progress can be made on the problem of automatic acquisition of knowledge. Of course, missing these particular points is a consequence of missing the existence of the AI problem as distinct from automation and study of the central nervous system.

3. A further disappointment is that chess playing programs have only reached an “experienced amateur” level of play. Well, if programs can’t do better than that by 1978, I shall lose a $L250$ bet and will be disappointed too though not extremely surprised. The present level of computer chess is based on the incorporation of certain intellectual mechanisms in the programs. Some improvement can be made by further refinement of the heuristics in the programs, but probably master level chess awaits the ability to put general configuration patterns into the programs in an easy and flexible way. I don’t see how to set a date by which this problem must be solved in order to avoid disappointment in the field of artificial intelligence as a whole.

4. Lighthill discusses the *combinatorial explosion* problem as though it were a relatively recent phenomenon that disappointed hopes that unguided theorem provers would be able to start from axioms representing knowledge

about the world and solve difficult problems. In fact, the *combinatorial explosion* problem has been recognized in AI from the beginning, and the usual meaning of *heuristic* is a device for reducing this explosion. Regrettably, some people were briefly over-optimistic about what general purpose heuristics for theorem proving could do in problem solving.

Did We Deserve It?

Lighthill had his shot at AI and missed, but this doesn't prove that everything in AI is ok. In my opinion, present AI research suffers from some major deficiencies apart from the fact that any scientists would achieve more if they were smarter and worked harder.

1. Much work in AI has the "look ma, no hands" disease. Someone programs a computer to do something no computer has done before and writes a paper pointing out that the computer did it. The paper is not directed to the identification and study of intellectual mechanisms and often contains no coherent account of how the program works at all. As an example, consider that the SIGART Newsletter prints the scores of the games in the ACM Computer Chess Tournament just as though the programs were human players and their innards were inaccessible. We need to know why one program missed the right move in a position - what was it thinking about all that time?

We also need an analysis of what class of positions the particular one belonged to and how a future program might recognize this class and play better.

2. A second disease is to work only on theories that can be expressed mathematically in the present state of knowledge. Mathematicians are often attracted to the artificial intelligence problem by its intrinsic interest. Unfortunately for the mathematicians, however, many plausible mathematical theories with good theorems such as control theory or statistical decision theory have turned out to have little relevance to AI. Even worse, the applicability of statistical decision theory to discriminating among classes of signals led to the mistaken identification of perception with discrimination rather than with description which so far has not led to much mathematics. More recently, however, problems of theorem proving and problems of representation have led to interesting mathematical problems in logic and mathematical theory of computation.

3. Every now and then, some AI scientist gets an idea for a general scheme of intelligent behavior that can be applied to any problem provided the machine is given the specific knowledge that a human has about the domain. Examples of this have included the GPS formalism, a simple predicate calculus formalism, and more recently the PLANNER formalism and perhaps the current Carnegie-

Mellon production formalism. In the first and third cases, the belief that any problem solving ability and knowledge could be fitted into the formalisms led to published predictions that computers would achieve certain levels of performance in certain time scales. If the inventors of the formalisms had been right about them, the goals might have been achieved, but regrettably they were mistaken. Such general purpose formalisms will be invented from time to time, and, most likely, one of them will eventually prove adequate. However, it would be a great relief to the rest of the workers in AI if the inventors of new general formalisms would express their hopes in a more guarded form than has sometimes been the case.

4. At present, there does not exist a comprehensive general review of AI that discusses all the main approaches and achievements and issues. Most likely, this is not merely because the field doesn't have a first rate reviewer at present, but because the field is confused about what these approaches and achievements and issues are. The production of such a review will therefore be a major creative work and not merely a work of scholarship.

5. While it is far beyond the scope of this review to try to summarize what has been accomplished in AI since Turing's 1950 paper, here is a five sentence try: Many approaches have been explored and tentatively rejected

including automaton models, random search, sequence extrapolation, and many others. Many heuristics have been developed for reducing various kinds of tree search; some of these are quite special to particular applications, but others are general. Much progress has been made in discovering how various kinds of information can be represented in the memory of a computer, but a fully general representation is not yet available. The problem of perception of speech and vision has been explored and recognition has been found feasible in many instances. A beginning has been made in understanding the semantics of natural language.

These accomplishments notwithstanding, I think that artificial intelligence research has so far been only moderately successful; its rate of solid progress is perhaps greater than most social sciences and less than many physical sciences. This is perhaps to be expected considering the difficulty of the problem.